Applications of Stochastics: Midterm

Some hints on the solutions

Oct 22, 2019

The absolute max is 40 points, but 35 points is already considered as 100%. You can write in Hungarian if that's much easier for you.

▷ **Exercise 1.** Let $X_k(n)$ be the number of degree k vertices in the Erdős-Rényi random graph $G(n, \lambda/n)$, with any $\lambda \in \mathbb{R}_+$ fixed. Show that $X_k(n)/n$ converges in probability, as $n \to \infty$, to $\mathbf{P}[\,\mathsf{Poisson}(\lambda) = k\,]$, for any $k = 0, 1, 2, \ldots$. [**8 points**]

Hints. Compute the first and second moments of $X_k(n)/n$, then use Chebyshev. For the first moment you will need the probability that a given vertex has degree k, which is obviously $\binom{n-1}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-1-k}$, though many people got it wrong.

▷ **Exercise 2.** Show that the Erdős-Rényi random graph $G(n, 3/n)$ contains some cycle (i.e., is not a forest) with probability tending to 1, as $n \to \infty$. [**8 points**]

Hints. If you try to do this by looking at the first and second moments of the total number of cycles, $N = \sum_{k=3}^{n} N_k$, where N_k is the number of k-cycles, it will be nightmarish, because you would need all the covariances between possible k-cycles and ℓ-cycles, for all possible pairs of values $3 \le k, \ell \le n$. Nevertheless, showing that $\mathbf{E}N \to \infty$ as $n \to \infty$ would already be a good first step. For this, you have to be careful not to give $\mathbf{E}N_k$ only up to constant factors that depend on k, because then you would have no idea how big $\sum_k$ is. So, notice that $\mathbf{E}N_k = \binom{n}{k}\frac{k!}{2k}\left(\frac{3}{n}\right)^k$, since there are $\frac{k!}{2k}$ ways to draw a k-cycle on k given points (check it for yourself). So, as $n \to \infty$, we have $\mathbf{E}N \sim \sum_{k=3}^{n}\frac{3^k}{2k}$, which goes to ∞ very fast.

But the real solution is much simpler; you just have to think a little bit instead of trying to repeat solutions to earlier exercises. Any forest on n vertices has at most $n-1$ edges. The number of edges in $G(n, 3/n)$ is $\mathsf{Binom}\left(\binom{n}{2}, \frac{3}{n}\right)$, with mean and variance both $\sim \frac{3}{2}n$, so Chebyshev gives that the graph has more than $n-1$ edges with probability tending to 1.

▷ **Exercise 3.** The apples from an orchard of a mathematician have random weights, distributed according to $\mathsf{Unif}[100, 300]$ grams, independently of each other. Johnny puts 100 apples in a box. But his doctor has forbidden him from lifting anything heavier than 25 kg. Give the best upper bound you can on the probability that the box is heavier than that. [**8 points**]

Hints. Please note that $X_1 + \cdots + X_{100}$, a sum of iid variables, has nothing to do with $100X_1$. For instance, the standard deviation of the first one is $\sqrt{100}\,\mathbb{D}(X_1)$, while it is $100\,\mathbb{D}(X_1)$ for the second one. When you want to estimate a sum of iid variables, Markov's inequality is a possibility, but is very weak, Chebyshev is better (assuming that the variable has finite variance), and exponential Markov is much better (assuming that the variable has a finite moment generating function). Giving an estimate via the Central Limit Theorem is also a possibility, but there you won't really know how the estimate compares to reality (is the sample 100 large enough? isn't the question $>$ 25 kg too extreme?), so probably exponential Markov is the best here (which, remember, is asymptotically optimal on the exponential scale, once you have optimized the value of t in e^{tX}).

▷ **Exercise 4.** Let $p, \alpha \in (0,1)$ arbitrary, and let $\alpha_n \to \alpha$ such that $\alpha_n n \in \mathbb{Z}$ for every n. Using Stirling's formula, find

$$\lim_{n\to\infty} \frac{-\log \mathbf{P}\big[\,\mathsf{Binom}(n,p) = \alpha_n n\,\big]}{n}.$$

[6 points]

Hints. From Stirling's formula for $n!$, if you don't mess up the calculations, it's quite straightforward.

▷ **Exercise 5.** Recall that the Cauchy distribution has density $\frac{1}{\pi(1+x^2)}$, and that if $X_1, \dots, X_n$ are i.i.d. Cauchy variables, then the sum $S_n = X_1 + \cdots + X_n$ has the distribution of nX_1. Prove the following:

(a) $S_n/n \xrightarrow{p} 0$ does not hold. **[2 points]**

(b) $S_n/n^{1.01} \xrightarrow{p} 0$. **[2 points]**

(c) $S_n/n^{2.01} \xrightarrow{\text{a.s.}} 0$. **[3 points]**

(d) For any $\epsilon > 0$, the expected number of returns to the interval $(-\epsilon, \epsilon)$ by the Cauchy walk S_n is infinite. **[3 points]**

Hints. We did (a) and (d) in class. For (b), fix $\epsilon > 0$, and notice that $\mathbf{P}\big[\,|S_n|/n^{1.01} > \epsilon\,\big] = \mathbf{P}\big[\,|X_1| > \epsilon n^{0.01}\,\big]$, which is an increasing sequence of events as $n \to \infty$, hence the limit of the probabilities is the probability of the union, which is just $\mathbf{P}[\,|X_1| = \infty\,] = 0$.

For (c), you first of all have to understand what almost sure convergence means. It is about the entire random sequence $(S_1/1,\ S_2/2^{2.01}, S_3/3^{2.01}, \dots)$, so the argument of (b) will not suffice here. You should apply the Borel-Cantelli lemma: if, for any $\epsilon > 0$, we have $\sum_{n\geq 1} \mathbf{P}[\,|S_n|/n^{2.01} > \epsilon\,] < \infty$, then $\{|S_n|/n^{2.01} > \epsilon\}$ happens only finitely often almost surely, hence we have the convergence to 0. Now you can already use that $\mathbf{P}[\,|S_n|/n^{2.01} > \epsilon\,] = \mathbf{P}[\,|X_1|/n^{1.01} > \epsilon\,] = 2\int_{\epsilon n^{1.01}}^{\infty} \frac{1}{\pi(1+x^2)} dx \asymp \frac{1}{\epsilon n^{1.01}}$, which is indeed summable in n.